

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 03:18:06 ; Search time 8.86557 Seconds  
(without alignments)  
1581.285 Million cell updates/sec

Title: US-09-807-933B-1

Perfect score: 1836  
Sequence: 1 MKFTTASSALLALALGTEM.....TYKEVTCPEKITAKTGCSSRK 338

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	694	37.8	376	1 GUNK_FUSOX	P45699 fusarium ox
2	679	37.0	213	1 GUNK_HOMIN	P43316 humicola in
3	552.5	30.1	333	1 GUNK_PSEFL	P18126 pseudomonas
4	515.5	28.1	393	1 GUNK_USTMA	P54424 ussillago ma
5	241	13.1	471	1 GUNK_TRIRE	P07987 trichoderma
6	180	9.8	418	1 GUNK_TRIRE	P07982 trichoderma
7	179	9.7	462	1 GUNK_FUSOX	P46236 fusarium ox
8	173	9.4	438	1 GUNK_AGABI	P49075 agaricus bi
9	153.5	8.4	385	1 GUNK_FUSOX	P46239 fusarium ox
10	150	8.2	2704	1 G168_PARP	P17053 paramedium
11	149.5	8.1	475	1 SIM1_YEAST	P40472 saccharomyc
12	148.5	8.1	540	1 GUNK_ASPAC	O59843 aspergillus
13	148	8.1	210	1 PSB_PORPU	P50272 porphyra pu
14	147.5	8.0	442	1 CYSA_DICDI	P54639 dictyostell
15	147	8.0	242	1 GUNK_TRIRE	P43317 trichoderma
16	144	7.8	2715	1 G156_PARP	P13397 paramedium
17	143	7.8	1396	1 VLTF_BPT5	P13390 bacterioph
18	141	7.7	962	1 GUNA_PSEFL	P10476 pseudomonas
19	138	7.5	485	1 Y136_TREPA	O83172 treponema p
20	138	7.5	506	1 GUNK_AGABI	O92400 agaricus bi
21	135.5	7.4	881	1 YJH8_YEAST	P47033 saccharomyc
22	135	7.4	388	1 GUNK_HOMIN	O12624 humicola in
23	133	7.2	513	1 GUNK_TRIRE	P00725 trichoderma
24	132	7.2	513	1 GUNK_TRIRE	P19355 trichoderma
25	130	7.1	677	1 SP87_DICDI	P54643 dictyostell
26	129.5	7.1	687	1 VS41_GIALA	P92127 giardia lam
27	129	7.0	316	1 LORI_HUMAN	P23490 homo sapien
28	128.5	6.9	389	1 SERI_BOMMO	P07856 bombyx mori
29	127.5	6.9	523	1 P60_LISSE	O01838 listeria se
30	127.5	6.9	537	1 SP70_DICDI	P15269 dictyostell
31	126	6.9	514	1 GUNK_FUSOX	P46239 fusarium ox
32	125	6.8	1218	1 JAG1_HUMAN	P78504 homo sapien
33	124.5	6.8	420	1 SUN4_YEAST	P53616 saccharomyc

34	124.5	6.8	459	1 GUNK_TRIRE	P07981 trichoderma
35	124	6.8	1218	1 JAG1_MOUSE	O9gx00 mus musculu
36	123.5	6.7	937	1 HYL1_CANAL	P46591 candida alb
37	123.5	6.7	1140	1 YMG6_YEAST	O04893 saccharomyc
38	122.5	6.7	450	1 UTH1_YEAST	P36135 saccharomyc
39	122.5	6.7	1210	1 ICEN_PSEFL	P09815 pseudomonas
40	121.5	6.6	503	1 YN23_YEAST	P53832 saccharomyc
41	121.5	6.6	604	1 AMYG_RHIO	P07683 rhizopus or
42	121	6.6	537	1 GUNK_PENNA	O06886 penicillium
43	121	6.6	1150	1 APNU_PIG	P12021 sus scrofa
44	120.5	6.6	1196	1 ICEV_PSEEX	O33479 pseudomonas
45	120	6.5	320	1 CEL1_AGABI	O00023 agaricus bi

## ALIGNMENTS

RESULT 1	ID	GUNK_FUSOX	STANDARD	PRT	376 AA.
AC	P45699				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	01-NOV-1995 (Rel. 32, Last annotation update)				
DE	Putative endoglucanase type K precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).				
OS	Fusarium oxysporum.				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;				
OX	Hypocreales; mitosporic Hypocreales; Fusarium.				
OX	NCBI_TaxID=5507;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=95047531; PubMed=7959045;				
RA	Sheppard P.O., Grant F.J., Oort P.J., Sprecher C.A., Foster D.C.,				
RA	Hagen F.S., Upshall A., McKnight G.L., O'Hara P.J.;				
RT	"The use of conserved cellulase family-specific sequences to clone				
RL	cellulase homologue cDNAs from Fusarium oxysporum.";				
RL	Gene 150:163-167(1994).				
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic				
CC	linkages in cellulose.				
CC	-1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).				
CC	-1- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL				
CC	HYDROLASES).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
CC	EMBL; L29381; AAA65589.1; -				
DR	HSSP; P43316; 2ENG.				
DR	InterPro; IPR000254; CBD_fungal.				
DR	InterPro; IPR000334; GH_45.				
DR	Pfam; PR00734; CBM_1; 1_				
DR	Pfam; PR02015; GLYCO_hydro_45; 1.				
DR	SMART; SM00236; ICBID; 1.				
DR	PROSITE; PS00562; CBD_FUNGAL; 1.				
DR	PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.				
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.				
FT	SIGNAL	1	18	POTENTIAL.	
FT	CHAIN	19	376	POTATIVE ENDOGLUCANASE TYPE K.	
FT	DOMAIN	19	308	CATALYTIC.	
FT	DOMAIN	309	338	LINKER.	
FT	DOMAIN	339	376	CELLULOSE-BINDING.	
FT	ACT SITE	29	29	NCLEOPHILE (BY SIMILARITY).	
FT	ACT SITE	140	140	PROTON DONOR (BY SIMILARITY).	
SQ	SEQUENCE	376 AA.	39235 MM;	B430A5F962B9F882 CRC64;	

Query Match 37.8%; Score 694; DB 1; Length 376;  
Best Local Similarity 58.6%; Pred. No. 3.2e-39;

Matches 126; Conservative 33; Mismatches 48; Indels 8; Gaps 5;

Query 124 AVSGAGSNGVTRRYDCCKACASGMPGKANVSPVSKCNKDVLTALSDNSAOGCNGNS 183  
 14 AVS-AASGSHSRMYDCCKPCSCSMGKAVALPALTCNDK-NPISNTNAVNGCGGGG 71  
 Db 184 -YACNDNQPAVNDNLAYGFAAALISGGESRWCCSCFELTFTSTSVAGKRVVQVNTNG 242  
 72 AYACTNTPMAVNDLAVGFAATKISGSEASWCCACTALTFTTGPKVKKMIVQSTNIG 131  
 QY 243 GDLGSSSTGAFDLOMPGCGVIGTNGCSSQWGAIPNDGWSRYGSISSASDCSSLPALQAG 302  
 132 GDLDGN---HFDLMPGCGVIGTNGCSSQWGAIPNDGWSRYGSISSASDCSSLPALQAG 186  
 Db 303 CKMRFNFPKADNPSTYKEVTCPEKITAKTCSR 337  
 187 CHMRFDFWENADNDFTEFQVQCPKALLDISGCKR 221

## RESULT 2

GUNS\_HUMIN STANDARD; PRT; 213 AA.

AC P4316;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Endoglucanase V (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)  
 DE (Cellulase V) (EG V).  
 OS Humicola insolens.  
 OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.  
 NCBI\_TaxID=34413;  
 RN [1] SEQUENCE FROM N.A.  
 RP Raemussen G., Mikkelsen J.-M., Schulein M., Patkar S.A., Hagen F.,  
 RA Hjort C.M., Hestrup S.;  
 RT "A cellulase preparation comprising an endoglucanase enzyme.";  
 RL Patent number WO9117243, 14-NOV-1991.  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).  
 RX MEDLINE=93390621; PubMed=8377880;  
 RA Davies G.J., Dodson G.G., Hubbard R.E., Tolley S.P., Dauter Z.,  
 RA Wilson K.S., Hjort C., Mikkelsen J.M., Raemussen G., Schulein M.;  
 RT "Structure and function of endoglucanase V.";  
 RL Nature 365:362-364 (1993).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RX MEDLINE=96101453; PubMed=8519779;  
 RA Davies G.J., Tolley S.P., Henriksen B., Hjort C., Schulein M.;  
 RT "Structures of oligosaccharide-bound forms of the endoglucanase V  
 from Humicola insolens at 1.9-A resolution.";  
 RL Biochemistry 34:16210-16220 (1995).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).  
 RA Davies G.J., Dodson G.G., Moore M.H., Tolley S.P., Dauter Z.,  
 RA Wilson K.S., Raemussen G., Schulein M.;  
 RT "Structure determination and refinement of the Humicola insolens  
 endoglucanase V at 1.5-A resolution.";  
 RL Acta Crystallogr. D 52:7-17 (1996).  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 linkages in cellulose.  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL  
 HYDROLASES).  
 CC PDB; 2ENG; 08-DEC-96.  
 CC PDB; 3ENG; 16-JUN-97.  
 CC PDB; 4ENG; 16-JUN-97.  
 DR InterPro: IPR000334; GH\_45.  
 DR Pfam: PF02015; Glyco\_hydro\_45; 1.  
 DR PROSITE: PS01140; GLYCOSYL\_HYDROL\_45; 1.  
 KM Cellulose degradation; Hydrolysis; Glycosidase; 3D-structure.  
 FT ACT\_SITE 10  
 FT ACT\_SITE 10 NUCLEOPHILE.  
 FT PROTON\_DONOR.  
 SQ SEQUENCE 213 AA; 22864 MW; 24334301BA5BC804 CRC64;

Query Match 37.0%; Score 679; DB 1; Length 213;  
 Best Local Similarity 59.1%; Pred. No. 1.8e-38;  
 Matches 123; Conservative 31; Mismatches 44; Indels 10; Gaps 5;

QY 132 NGVTRRYDCCKACASGMPGKANVSPVSKCNKDVLTALSDNSAOGCNGNS 190  
 2 DGRSTRYDCCKPCSCSMGKAVALPALTCNDK-NPISNTNAVNGCGGGG 60  
 Db 191 PAVVNDNLAYGFAAALISGGESRWCCSCFELTFTSTSVAGKRVVQVNTNGDLSSTG 250  
 61 PAVVNDNLAYGFAAALISGGESRWCCSCFELTFTSTSVAGKRVVQVNTNGDLSSTG 118  
 QY 251 AHFDLOMPGCGVIGTNGCSSQWGAIPNDGWSRYGSISSASDCSSLPALQAGCKMRFN 309  
 119 -HFDLMPGCGVIGTNGCSSQWGAIPNDGWSRYGSISSASDCSSLPALQAGCKMRFN 173  
 QY 310 FKNADNPSTYKEVTCPEKITAKTCSR 337  
 174 FKNADNPSTYKEVTCPEKITAKTCSR 201

## RESULT 3

GUNS\_PSEFL STANDARD; PRT; 511 AA.

AC P18126;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Cellulase) (ECB).  
 OS CELB.  
 OS Pseudomonas fluorescens.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 NCBI\_TaxID=294;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-48.  
 RC STRAIN=SP. Cellulosa;  
 RX MEDLINE=90355836; PubMed=2117693;  
 RA Gilbert H.J., Hall J., Hazlewood G.P., Ferreira L.M.A.;  
 RT "The N-terminal region of an endoglucanase from Pseudomonas  
 fluorescens subspecies cellulosa constitutes a cellulose-binding  
 domain that is distinct from the catalytic centre.";  
 RL Mol. Microbiol. 4:759-767 (1990).  
 CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-  
 GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-  
 GLUCANS. ECB IS MOST ACTIVE AGAINST AMORPHOUS AND CRYSTALLINE CELLULOSE.  
 CC SIGNIFICANT ACTIVITY AGAINST AMORPHOUS AND CRYSTALLINE CELLULOSE.  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 linkages in cellulose.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN  
 (CBD).  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL  
 HYDROLASES).  
 CC EMBL; X52615; CA36844.1; -.  
 CC PIR; S10527; S10527.  
 DR HSSP; P43316; 2ENG.  
 DR InterPro: IPR001919; Bac cellose-bind.  
 DR InterPro: IPR002883; CBD\_5.  
 DR InterPro: IPR000334; GH\_45.  
 DR Pfam; PF00553; CBM\_2; 1.  
 DR Pfam; PF02013; CBM\_10; 1.  
 DR Pfam; PF02015; Glyco\_hydro\_45; 1.

DR PROSITE; PS00561; CBD BACTERIAL; 1.  
 DR PROSITE; PS01140; GLYCOSYL HYDROL P45; 1.  
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Periplasmic.  
 FT SIGNAL 1 23  
 FT CHAIN 30 511  
 FT DOMAIN 30 131  
 FT DOMAIN 132 173  
 FT DOMAIN 223 259  
 FT DISULFID 32 127  
 FT ACT SITE 276 276  
 FT ACT SITE 393 393  
 SQ SEQUENCE 511 AA; 52078 MW; 3C3119D98291D8E CRC64;

Query Match 30.1%; Score 552.5; DB 1; Length 511;  
 Best Local Similarity 35.2%; Pred. No. 9,4e-30;  
 Matches 146; Conservative 48; Mismatches 128; Indels 93; Gaps 16;

QY 4 ITTASALLALALGTEMASAECSKLYQ-CGKKNMNGPTCES-----GSTCKVAN 54  
 DB 98 IPIGSSVERGVQ-GNNGSPRAQVPAVTGACGGSSAPESVSSSSSVSTPRSSS 156  
 QY 55 DYVSQCLPSSSSGNKSESNAHK-----TTTAAHKTTT-----AAHKTTTAP 98  
 DB 157 SSVSSSVPTSSSSSVTLGAQACWYGTLPICNNTSNGWGYEDGRSCVARTTCSAOP 216  
 QY 99 AKKTTTAAKASTPSNSS-----SSGKYSAVSGASG-----NGVTRRYMD 140  
 DB 217 APYGIYSTSTSTLSSSSSSSSSVASSSSLSATSSASVSVPPIDGCKNGATRYMD 276  
 QY 141 CCKASCSWPGKANVS---SPVKSCKNDGVTALSDSNAQSGNGNSYMCNDNPMAVNDN 197  
 DB 277 CCKPHCGM--SANNVPLSVPLQSCSANN--TRLSDVSVGSSCDGGGVMCMCKI.PPAVSP 333  
 QY 198 LAYGFAAAISGGGSEBRCSCPELTFT-----STSVAGKKNVQVNTTGGDJA 246  
 DB 334 LAYGVAAT--SGQ-----VCCRCYLOFTGSSSYNAPGPPGSAALAGKMTIVATNIGYDV- 387  
 QY 247 SSTGAFHFDLQMPGGVGIFPGSSSSOMGAPNDGMSRGYGISA-----289  
 DB 388 --SGGQFDLVPGGGAGAFACSAQWVSNAELQAGGFLACCKQOLGYNASLSQYKSC 445  
 QY 290 --SDCSSL-----PSALQAGCKRFFWFKNADNPMTYKEVTCPKEITTAQTGCR 337  
 DB 446 VLMRCDSVFESRGLTQDQGCCTPAEFMEFAADNPULKYKEVPPAELTTRSGNMR 500

## RESULT 4

ID GUNI USTWA STANDARD; PRT; 393 AA.  
 AC P54424;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Endoglucanase 1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase 1)  
 DE (Cellulase 1) (EG 1).  
 GN EGL1.  
 OS Ustilago maydis (Smut fungus).  
 OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;  
 OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.  
 ON NCBI\_TaxID=5270;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PED11;  
 RX MEDLINE=96145728; PubMed=8590631;  
 RA Schamwecker F., Wanner G., Kahmann R.;  
 RT "Filament-specific expression of a cellulase gene in the dimorphic  
 fungus Ustilago maydis";  
 RL Biol. Chem. Hoppe-Seyler 376:617-625 (1995).  
 CC -1- CATALYTIC ACTIVITY: Endohydrolase of 1,4-beta-D-glucosidic  
 CC linkages in cellulose.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: HYPMAL TIP.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN FILAMENTOUS DIKARYON.

CC -1- PTM: MAY ALSO BE O-GLYCOSYLATED.  
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL  
 CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; S81598; AAB36147.1; -

DR HSSP; P43316; ZENG.

DR InterPro; IPR000334; GH 45.

DR Pfam; PF02015; Glyco\_hydro\_45; 1.

DR PROSITE; PS01140; GLYCOSYL HYDROL P45; 1.  
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Glycoprotein.

FT SIGNAL 1 26  
 FT CHAIN 27 393  
 FT ACT SITE 34 34  
 FT ACT SITE 152 152  
 FT ACT SITE 270 385  
 FT DOMAIN 343 343  
 FT CARBOHYD 343 343  
 SQ SEQUENCE 393 AA; 39594 MW; 65C753C610CD6AD3 CRC64;

Query Match 28.1%; Score 515.5; DB 1; Length 393;  
 Best Local Similarity 47.0%; Pred. No. 2e-27; Indels 21; Gaps 9;  
 Matches 103; Conservative 31; Mismatches 64;

QY 133 GVTTRYWDCCKASCSGPGKANVSPPVKSCKNDGVTAL---SDSNAQSGNGNSYMCNDN 189  
 DB 27 GMATRYWDCCLASASMEGAPYAPVADCKAGVTLIDSKDPSSGSGNGNKKPFCSGM 86  
 QY 190 QPW--AVNDNLAYGPAAAIISGGGSRMCCSCPELTFT---TSTSVAGKKNVQVNTTGG 243  
 DB 87 QPFDDEDTPLAFLGF--GAFITGGQSDTDCAFYAEFEHDAQKAMKRKLI.FQVYNNVG 144  
 QY 244 DLGSSGAFHFDLQMPGGVGIF-NGCSSQWGA.PNDGMSRGYGISASPSLQNG 302  
 DB 145 DVQSQ---NFDQIPGGGIGAFPKGCPAPQGVASLWGDQGVASVATSCSLPRLQRC 201  
 QY 303 CKMRFN-WFKNADNPMSM--TYKEVTCPKEITTAQTGCSRK 338  
 DB 202 CKMRSEW---GDNFVLKGSPPKRVKCPKSLIDRSQCRK 237

## RESULT 5

ID GUX2 TRIRE STANDARD; PRT; 471 AA.  
 AC P07987;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Exoglucanase II precursor (EC 3.2.1.91) (Exocellulohydrolase II)  
 DE (CBHII) (1,4-beta-cellulohydrolase).  
 GN CBH2.  
 OS Trichoderma reesei (Hypocrea jecorina).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreales; Hypocreaceae; Hypocrea.  
 ON NCBI\_TaxID=51453;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VT-D-80133;  
 RX MEDLINE=87248061; PubMed=3596237;  
 RA Teeri T.T., Lehtovaara P., Kauppinen S., Salovuori I., Knowles J.;  
 RT "Homologous domains in Trichoderma reesei cellulolytic enzymes: gene  
 RT sequence and expression of cellobiohydrolase II";  
 RL Gene 51:43-52 (1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=QM9414 / Rut C-30;  
 RA Chen C.M., Gritzalis M., Stafford D.W.;

RT "Nucleotide sequence and deduced primary structure of  
 RT cellobiohydrolase II from *Trichoderma reesei*.";   
 RL Biotechnology 5:274-278(1987).   
 RN [3]   
 RP SEQUENCE OF 25-44.   
 RA Faergestad L.G., Pettersson L.G.;   
 RT "The 1,4-beta-glucan cellobiohydrolases of *Trichoderma reesei* QM   
 RT 9414.";   
 RL FEBS Lett. 119:97-100(1980).   
 RN [4]   
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTOMS).   
 RX MEDLINE=9033255; Pubmed=2377893;   
 RA Rouvinen J., Bergfors T., Teeri T.T., Knowles J.K.C., Jones T.A.;   
 RT "Three-dimensional structure of cellobiohydrolase II from *Trichoderma   
 RT reesei*.";   
 RL Science 249:380-386(1990).   
 RN [5]   
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTOMS).   
 RX MEDLINE=97029636; Pubmed=8875646;   
 RA Kotivola A., Reihlakainen T., Ruohonen L., Valkenjaervi A.,   
 RA Claessens M., Telemann O., Kleywegt G.J., Szardenans M., Rouvinen J.,   
 RA Jones T.A., Teeri T.T.;   
 RT "The active site of *Trichoderma reesei* cellobiohydrolase II: the role   
 RT of tyrosine 169.";   
 RL Protein Eng. 9:691-699(1996).   
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE .   
 CC GENERALY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:   
 CC (1) ENDOCULICANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;   
 CC (2) EXOCULICOHYDROLASES THAT CUT THE DISACCHARIDE CELLULOSE   
 CC FROM THE NONREDUCTING END OF THE CELLULOSE POLYMER CHAIN:   
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER   
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.   
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-Delta-D-glucosidic linkages   
 CC in cellulose and cellobiose, releasing cellobiose from the non-   
 CC reducing ends of the chains.   
 CC -1- SUBCELLULAR LOCATION: Secreted.   
 CC -1- MISCELLANEOUS: T.REESEI PRODUCES TWO DIFFERENT   
 CC EXOCULICOHYDROLASES. THEY ARE UNIQUE IN THAT THEY CAN HYDROLYSE   
 CC CRYSTALLINE CELLULOSE IN THE ABSENCE OF ENDOGLUCANASES.   
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL   
 CC HYDROLASES).   
 CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).   
 CC -----   
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration   
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation --   
 CC the European Bioinformatics Institute. There are no restrictions on its   
 CC use by non-profit institutions as long as its content is in no way   
 CC modified and this statement is not removed. Usage by and for commercial   
 CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)   
 CC or send an email to [license@ib-sib.ch](mailto:license@ib-sib.ch)).   
 CC -----   
 DR EMBL, M16190, AAA34210.1; -.   
 DR EMBL, M55080, AAA72922.1; -.   
 DR PIR, A26472; A26472.   
 DR PIR, A26160; A26160.   
 DR PDB, 3CBH; 15-JAN-91.   
 DR PDB, 1CB2; 19-MAR-99.   
 DR InterPro, IPR000254; CBD\_fungal.   
 DR InterPro, IPR001524; GH\_6.   
 DR Pfam, PF00734; CBM\_1; 1.   
 DR Pfam, PF01341; Glyco\_hydro\_6; 1.   
 DR PRINTS, PR00733; GLHYDRASE6.   
 DR ProDom, PD001821; CBD\_fungal; 1.   
 DR ProDom, PD003733; GH\_6; 1.   
 DR SMART, SM00236; ECBD\_1.   
 DR PROSITE, PS00562; CBD\_FUNGAL; 1.   
 DR PROSITE, PS00655; GLYCOSYL\_HYDROL\_F6\_1; 1.   
 DR PROSITE, PS00656; GLYCOSYL\_HYDROL\_F6\_2; 1.   
 KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal;   
 KW 3D-structure.   
 FT SIGNAL 1 24   
 FT CHAIN 25 471   
 FT DOMAIN 25 65   
 CC EXOGLUCANASE II.   
 CC CELLULOSE-BINDING (BY SIMILARITY).

FT	DOMAIN	66	471	LINKER.
FT	DNAI1	107	471	CATALYTIC.
FT	ACT_SITE	199	199	PROTON DONOR.
FT	ACT_SITE	245	245	NUCLEOPHILE.
FT	ACT_SITE	425	425	BY SIMILARITY.
FT	DISULFID	34	51	BY SIMILARITY.
FT	DISULFID	45	61	BY SIMILARITY.
FT	DISULFID	200	259	
FT	DISULFID	392	439	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	38	38	O-LINKED (MAN. .)
FT	CARBOHYD	111	111	O-LINKED (MAN. .)
FT	CARBOHYD	121	121	O-LINKED (MAN. .)
FT	CARBOHYD	130	130	O-LINKED (MAN. .)
FT	CARBOHYD	133	133	O-LINKED (MAN. .)
FT	CARBOHYD	134	134	O-LINKED (MAN. .)
FT	CARBOHYD	139	139	O-LINKED (MAN. .)
FT	CARBOHYD	313	313	N-LINKED (GLCNAC. .)
FT	CARBOHYD	334	334	N-LINKED (GLCNAC. .)
FT	MUTASEN	199	199	D->A. 2% OF WILD-TYPE ACTIVITY.
FT	MUTASEN	245	245	P->A. NO MEASURABLE ACTIVITY.
FT	CONFLICT	359	359	P -> R (IN REF. 2).
FT	CONFLICT	449	449	P -> A (IN REF. 2).
SQ	SEQUENCE	471 AA;	49653 MW;	C471BDC35B1BD88 CRC64;
Query Match		13.1%;	Score 241;	DB 1; Length 471;
Best Local Similarity		40.5%;	Pred. No. 3.2e-05;	
Matches		51; Conservative	21; Mismatches	46; Indels 8; Gaps 3
Db	5 TIASSALTLALGTMASAAECSKLYGCGGGKMNNGPTCCSGSTCKVSNDDYYSOCLPSG	64		
Db	8 TLATLTLTLLASPLDERQA--CSSWVGCGGGGNMSPGCCAGSSTCYNSDYYSOCLPGA	65		
QY	65 SSGNSSSESAHKTTTAHHKTTTPAKHTTTAPAKTTTYAKASTPSNSSSSSGKYSA	124		
Db	66 ASSSSSTRRA---SITSRVSEFTTSRSSASPPEGSTTRVPVG--SGTATYSGNPFVG	119		
QY	125 VSGGAS	130		
Db	120 VTIPWAN	125		
RESULT 6				
GUN2_TRIRE				
ID	_GUN2_TRIRE	STANDARD;	PRT;	418 AA.
AC	P07983;			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-AUG-1988 (Rel. 08, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Endoglucanase EG-II precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (cellulase).			
GN	EGII OR EGI.II.			
OS	Trichoderma reesei (Hypocrea jecorina).			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
OC	Hypocreales; Hypocraeceae; Hypocrea.			
NCBI_TaxID=51453;				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=VT-D-80133;			
RX	MEDLINE=8825850; PubMed=3384334;			
RA	Johanson M., Lehtoavaara P., Penttilä M., Teeri T.T., Staahlberg J.,			
RT	"EdIII, a new endoglucanase from Trichoderma reesei: the			
RL	characterization of both gene and enzyme."			
RL	Gene 63:11-21(1988).			
RN	[2]			
RP	ACTIVE SITE GLU-350.			
RA	MEDLINE=93131031; PubMed=8093602;			
RT	Macarron R., van Beumen J., Henriksat B., de la Mata I.,			
RT	Clayessens M.;			
RT	"Identification of an essential glutamate residue in the active site			
RT	of endoglucanase III from Trichoderma reesei";			
CC	-FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE			



[illegible]



```

DR InterPro: IPR001000; Glyco_hydro.10.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR Pfam: PF00734; CBM_1.1.1.
DR PRINTS: PRO0134; GLYHYDRLASE10.
DR ProDom: PD001821; CBM_Fungal; 1.
DR SMART: SM00236; fCBM; 1.
DR PROSITE: PS00562; CBM_FUNGAL; 1.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
KM Cellulose degradation; Hydrolyase; Glycosidase; Signal.
FT SIGNAL 1 19
FT CHAIN 20 385
FT DOMAIN 25 52
FT DOMAIN 53 84
FT DOMAIN 85 385
FT ACT SITE 209 209
FT ACT SITE 321 321
SQ SEQUENCE 385 AA; 41225 MW; B3C3807C0D73C0EC CRC64;

Query Match 8.4%; Score 153.5; DB 1; Length 385;
Best Local Similarity 25.0%; Pred. No. 0.0016;
Matches 46; Conservative 20; Mismatches 59; Indels 59; Gaps 6;

QY 9 SALLALALGTEMAAECSEKLYGCGKMMNGPTCCSGSTKYNDYSSQCLPSSGN 68
   ||||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 SVLLALAPVSALAQA-----PIWGCGGNGMTGATTCASGLKCEKINDWYQCVF-GSGG- 58

QY 69 KSESASHKTTTAAHKTTTAAHKTTTAAKTITVAKASTPSSSSSSSGKYSVSGG 128
   -----SEPCPSSTGGGTPPTGPN 78

Db 129 ASGNGVTRTYWDCCKASCSPGKANYSPVSKCNKQDVTALSDSNNGSGNGSNYMCND 188
   : : : : : : : : : : : : : : : : : : : : : : : :
QY 79 SGGTGLDAK-----KAK-----GKQYFGTEIDHHLNPLINIVKQFG-----QVTCTN 125

QY 189 NQFW 192
   :
Db 126 SMKW 129

RESULT 10
G168_PARP STANDARD; PRT; 2704 AA.
AC P17053;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE 168G surface protein precursor.
GN 168G.
OS Parametium primaurelia.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Parametium.
OX NCBI_TaxID=5886;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=90172419; PubMed=2308165;
RA Prar A.;
RT "Conserved sequences flank variable tandem repeats in two alleles of
   the G surface protein of Parametium primaurelia.";
RL J. Mol. Biol. 211:521-535(1990).
CC -1- FUNCTION: THIS PROTEIN IS THE SURFACE ANTIGEN OR IMMOBILIZATION
   ANTIGEN OF PARAMETIUM PRIMAURELIA.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- MISCELLANEOUS: IT HAS INTERNAL HOMOLOGIES AND A HIGHLY PERIODIC
   STRUCTURE WITH 37 PERIODS OF ABOUT 75 RESIDUES, EACH PERIOD
   CONTAINING 8 CYSTEINES, EXCEPT FOR FOUR HALF PERIODS. A VARIABLE
   PART OF 475 RESIDUES COMPRISES 4 ALMOST IDENTICAL PERIODS IN THE
   MIDDLE OF THE PROTEIN.
CC -1- MISCELLANEOUS: EXPRESSION OF G PROTEIN OCCURS AT LOW TEMPERATURES
   (14-32 DEGREES CELSIUS).
CC -1- SIMILARITY: 98% TO THE ALLELIC FORM 156G PROTEIN (AC P13837) IN
   PERIODIC STRUCTURE AND 80% IN THE VARIABLE DOMAIN IN THE MIDDLE
   OF THE PROTEIN.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X52133; CAA36378.1; -.
DR PIR: S09118; S09118.
DR InterPro: IPR002895; Parametium SA.
DR Pfam: PF01508; Parametium SA; 34.
KM Signal; Repeat; Antigen; Membrane; GPI-anchor.
FT SIGNAL 1 20
FT CHAIN 21 2704
FT DOMAIN 106 2560
FT DOMAIN 1060 1424
SQ SEQUENCE 2704 AA; 278775 MW; 40EAD0AB18BE2119 CRC64;

Query Match 8.2%; Score 150; DB 1; Length 2704;
Best Local Similarity 22.3%; Pred. No. 0.017;
Matches 82; Conservative 33; Mismatches 148; Indels 104; Gaps 17;

QY 30 YGCGGKMMNGPTCC-----SGSTC-----KYSND-----YSSQCLPSSGN 68
   ||||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 2271 YTSCKSLMNMNDSCMWISNCKTTGNSNCVGTLCSETITDGGCVTGDGACIQSVPALN 2330

QY 69 KSESASHKTTTAAHKTTTAAHKTTTAAKTITVAKASTPSSSSSSSGKYSVSGG 124
   -----SEPCPSSTGGGTPPTGPN 78

Db 2331 SSDPKVKRYTSCDAFYTT--HSDQQLASSKCTTNGTGIALSSCSYTAQAGCYFND 2388

QY 125 -----VSGASGNGVTRTYWDCCKASC-----SWEKANYSPVSKCNKQDVTAL-- 169
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 2389 KGTLYTSGVITSTGICT--WDTTSSSCRDQSCADLTGTHATCSSLQSLSTGSDGTCLLK 2446

QY 170 -----SDSNNGSGNG--GNSYMCNDNQPMAYVNDNLAYG--FAAAVSGGESRMC----- 216
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 2447 GACTSYTTDTACTTAVGSDGACYWELASRTNNNTAKRLTLADIONGTATNVCSVALST 2506

QY 217 -----CSCF--ELFTSTSVAGKMMVQVNTVGGDLGSGTGAHFDLQPMGGG 261
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 2507 CVSNGTACIPKANCSTYTSKINCNSGGDLGICVFPQSTATGAAGTGCTA----- 2556

QY 262 VGIPIFGCSSQWAPND-----GW--GSRYGISASDCS---SLPSALQAGCK 304
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 2557 --LMTACTV---ANNQDQCAARDRCGWTASGTGATVASKCATHTCATNQTATGACT 2611

QY 305 WRFNWF 311
   ||
Db 2612 RFLNWDK 2618

RESULT 11
SIM1_YEAST STANDARD; PRT; 475 AA.
AC P40472;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE SIM1 protein precursor.
GN SIM1 OR PB3 OR YIL123W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
   Gentles S., Hamlyn N., Horenell T.S., Hunt S., Jagsels K., Jones M.,

```

RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,  
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
 RA Walsh S.V., Whitehead S.,  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP IDENTIFICATION.  
 RA MEDLINE=86155611; PubMed=8574583;  
 RA Dahmann C., Diffley J.F.X., Nasmyth K.A.;  
 RT "S-phase-promoting cyclin-dependent kinases prevent re-replication by  
 RT inhibiting the transition of replication origins to a pre-replicative  
 RT state.";  
 RL Curr. Biol. 5:1257-1269 (1995).  
 CC -1- FUNCTION: REQUIRED FOR THE MAINTENANCE OF THE CLB5 KINASE  
 CC ACTIVITY.  
 CC -1- SIMILARITY: BELONGS TO THE SUN FAMILY.  
 CC -1- ACTIVITY: BELONGS TO THE SUN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z46833; CAA86869.1; -  
 DR SGD: S0001385; SIM1.  
 KW Signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 475 SIM1 PROTEIN.  
 FT DOMAIN 58 112 ALA-RICH.  
 FT DOMAIN 80 203 SER-RICH.  
 FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 475 AA; 48070 MW; B7B0F0BC8478612P CRC64;  
 Query Match 8.1%; Score 149.5; DB 1; Length 475;  
 Best Local Similarity 20.4%; Pred. No. 0.0035;  
 Matches 71; Conservative 47; Mismatches 131; Indels 99; Gaps 13;  
 Oy 5 TTASALLALGTMASAGSKLYGCGGKMMNPFCCSGSTCKVSNQDYSCLPSPG 64  
 Db 79 TSSAAGTAASINAVSALLAKREKISDAAA---SATSTSGGASSSSSSSSASTLESS 134  
 Oy 65 SSGNKSSESAAHKT---TTAAHKKTTTAAHKKTTTAAKSTTAAKASTPSSSSSSSGK 121  
 Db 135 SVSSSSSEAAPTSTVSTTSATQSSASATKSTSTSTSTSTSTSTSTSTSTSSSSSSSS 194  
 Oy 122 YSANGGA---SGNGYTRRYV----- 139  
 Db 195 SSSSSSGSISYGLDLPFGSPSEKFPDGTTPCDKFPSPGGGVISIDWIGGMSGVENDTDS 254  
 Oy 140 ---DC-----CKASGMPGKANVSPVKSCKNDGVTALSDSNAQSG--CNGNSYMCNDN 189  
 Db 255 TGSGCKESGYSYSCQ-PGMSKTQMPSDO-----PSDGRSVGGLCKRGYLYRSTTD 305  
 Oy 190 QPMVAVNDNLVAFPAALASGGESRWCCSCEFLFTSTSVAG-KKVVQVNTVNGDL--- 245  
 Db 306 ADYLCEWGEAAVYVSKLSKG-----VAICRPDYPTENNVIPYVEGGSSHPL 354  
 Oy 246 -----GSSTGAHFDLQMPGGGVGFNGSSSQMGAPNDWMS 281  
 Db 355 TVVDODVFTWEGKTSAY--YNNAGVSVEDGCI--WGTSGGSIGN 398  
 RESULT 12  
 GUXI\_ASPAC STANDARD; PRT; 540 AA.  
 ID GUXI\_ASPAC  
 AC 059843;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase I) (1,4-  
 DE beta-cellulohydrolase I) (beta-glucanancellobiohydrolase I).  
 GN CBHI.

OS Aspergillus aculeatus.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eucristales; Trichocommataceae; mitosporic Trichocommataceae; Aspergillus.  
 OK NCBI\_TaxID=5053;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=F-50;  
 RA Takada G., Kawaguchi T., Sumitani J., Arai M.;  
 RT "Cloning, nucleotide sequence, and transcriptional analysis of  
 RT Aspergillus aculeatus No.F-50 cellobiohydrolase I (cbhi) gene.";  
 RL J. Ferment. Bioeng. 85:1-9 (1998)  
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE  
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:  
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;  
 CC (2) EXOCELLULOBIODOLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE  
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;  
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYSE THE CELLOBIOSE AND OTHER  
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages  
 CC in cellulose and cellobiose, releasing cellobiose from the non-  
 CC reducing ends of the chains.  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL  
 CC HYDROLASES).  
 CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB002821; BAA25183.1; -  
 DR HSSP: P00725; IAZ6.  
 DR InterPro: IPR000254; CBD\_fungal.  
 DR InterPro: IPR001722; GH\_7.  
 DR Pfam: PF00734; CBM\_1; 1.  
 DR Pfam: PF00840; Glyco\_hydro\_7; 1.  
 DR Prodom: PD001821; CBD\_fungal; 1.  
 DR Prodom: PD186135; GH\_7; 1.  
 DR SMART: SM00236; fcbd\_1.  
 DR PROSITE: PS00562; CBD\_FUNGAL; FALSE NEG.  
 DR Cellulose degradation; Hydrolyase; Glycosidase; Glycoprotein; Signal.  
 KW SIGNAL 1 22  
 FT CHAIN 23 540  
 FT DOMAIN 23 540 EXOGLUCANASE I.  
 FT DOMAIN 23 540 CATALYTIC.  
 FT DOMAIN 506 540 LINKER.  
 FT ACT\_SITE 234 234 CELLULOSE-BINDING (BY SIMILARITY).  
 FT ACT\_SITE 239 239 PROTON DONOR (BY SIMILARITY).  
 FT DISULFID 513 529 BY SIMILARITY.  
 FT DISULFID 524 539 BY SIMILARITY.  
 FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 540 AA; 57099 MW; 002D7FD28D194D0 CRC64;  
 Query Match 8.1%; Score 148.5; DB 1; Length 540;  
 Best Local Similarity 59.1%; Pred. No. 0.0046;  
 Matches 26; Conservative 5; Mismatches 12; Indels 1; Gaps 1;  
 Oy 18 TENASAAECSKLYGCGGKMMNPFCCSGSTCKVSNQDYSCQL 61  
 Db 498 TTSSSTTNVVALYGGCGGGMGTGTCASG-TCKNDYISQCL 540  
 RESULT 13  
 PSBP\_PORPU STANDARD; PRT; 210 AA.  
 ID PSBP\_PORPU  
 AC PS0272;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Putative polysaccharide binding protein precursor.  
 OC Porphyra purpurea.  
 RC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.  
 OK NCBI\_TaxID=2787;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Avonport;  
 RA Liu O., der Meer J.P., Reith M.E.;  
 RL Submitted (May-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 4.FUNGAL-TYPE CELLULOSE-BINDING DOMAINS (CBD).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U08843; AAA61792.1; -  
 CC HSSP; P00725; 2CBH.  
 CC InterPro; IPR000254; CBD\_fungal.  
 CC Pfam; PF00734; CBM\_1; 4.  
 CC SMART; SM00236; FCBF; 4.  
 CC PROSITE; PS00562; CBD\_FUNGAL; 4.  
 CC Repeat; Signal; 1.  
 CC SIGNAL 1 22  
 CC CHAIN 23 210 PUTATIVE POLYSACCHARIDE BINDING PROTEIN.  
 CC DOMAIN 23 62 POLYSACCHARIDE BINDING DOMAIN I.  
 CC DOMAIN 63 105 POLYSACCHARIDE BINDING DOMAIN II.  
 CC DOMAIN 125 165 POLYSACCHARIDE BINDING DOMAIN III.  
 CC DOMAIN 166 210 POLYSACCHARIDE BINDING DOMAIN IV.  
 CC SEQUENCE 210 AA; 21967 MW; 83782D98B683CC8 CRC64;  
 Query Match 8.1%; Score 148; DB 1; Length 210;  
 Best Local Similarity 23.9%; Pred. No. 0.002;  
 Matches 56; Conservative 17; Mismatches 82; Indels 86; Gaps 10;  
 QY 13 ALALGTMSAESAECISKLYGCGGKNNNGPTCCSGSTCKYSNDYSSQC-----LP 62  
 DB 10 ALTLTSLAALAAASACGLVYEQCGIGFPGVTCSEGLCMKMGKPYSSCRAMPGMGQVKP 69  
 QY 63 SGSSGKSSBSAHKKTTHAHHKTTTAHKKTTTAAKKTTHV-----AKAST 110  
 DB 70 YGCGGKNNYSG-----KTMCSPEPKVELNEFPSSQCDLANKSPVAT 110  
 QY 111 PSNSSSSSS-----SGKTSAVSGGASGNGVTRRYMDCK-----ASCMPGKANYSSP 157  
 DB 111 PKVSPSPPPAPQVCGKEVLAACGEMPMGA-----KCKRGLVCEYETSGMKMGSCRAAPP 165  
 QY 158 VKSCNKDGVTLSDSNAGSGNGNSY-----MCDNDNPMAVNDNLAYGFAAAAISSGGES 213  
 DB 166 -----KMGVGRYAQC-GGMGYMGSTMC-----VGGYKCMALISGSMY 202  
 QY 214 RMC 216  
 DB 203 KQC 205  
 RESULT 14  
 ID CYS4\_DICDI STANDARD; PRT; 442 AA.  
 AC P54639;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cysteine proteinase 4 precursor (EC 3.4.22.-).  
 GN CPD OR CP4.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.  
 OK NCBI\_TaxID=44689;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=AX4;  
 RX MEDLINE=96081966; PubMed=7499424;  
 RA Souza G.M., Hirai J., Mehta D.P., Freeze H.H.;  
 RT "Identification of two novel Dictyostelium discoideum cysteine  
 RT proteinases that carry N-acetylglucosamine-1-P-modification";  
 RL J. Biol. Chem. 270:28938-28945 (1995).  
 CC -1- SUBCELLULAR LOCATION: Lysosomal.  
 CC -1- DEVELOPMENTAL STAGE: PRESENT IN THE VEGETATIVE PHASE AND DECREASES  
 CC WITH THE START DEVELOPMENT, REAPPEARS IN LOW LEVELS WHEN THE  
 CC FRUITING BODY IS FORMED.  
 CC -1- PTM: PHOSPHOGLYCOSYLATED, CONTAINS GLCNAC-ALPHA-1-P-SER RESIDUES  
 CC AND FUCCSE.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L36204; AAA92019.1; -  
 CC HSSP; P07711; 1CUL.  
 CC MEROPS; C01.081; -  
 CC DictyDb; DD01060; CPD.  
 CC InterPro; IPR000668; Peptidase\_C1.  
 CC InterPro; IPR00169; SHprot\_acsite.  
 CC Pfam; PF00112; Peptidase\_C1; 1.  
 CC PRINTS; PR00705; PAPAIN.  
 CC PRODOM; PD000158; Peptidase\_C1; 2.  
 CC DR PROSITE; PS00639; TH1OL\_PROTEASE\_CYS; 1.  
 CC DR PROSITE; PS00640; TH1OL\_PROTEASE\_HIS; 1.  
 CC DR PROSITE; PS00640; TH1OL\_PROTEASE\_ASN; 1.  
 CC KM Hydroxylase; Thiol protease; Lysozyme; zymogen; glycoprotein;  
 CC Phosphorylation; Signal; Repeat.  
 CC SIGNAL 1 17  
 CC PROPR 18 111  
 CC CHAIN 112 442  
 CC DOMAIN 286 296  
 CC DOMAIN 299 304  
 CC DOMAIN 310 315  
 CC DOMAIN 319 323  
 CC DOMAIN 327 332  
 CC ACT\_SITE 135 135  
 CC ACT\_SITE 277 277  
 CC ACT\_SITE 406 406  
 CC DISULFID 132 178  
 CC DISULFID 169 212  
 CC DISULFID 270 428  
 CC CARBOHYD 228 228  
 CC CARBOHYD 254 254  
 CC SEQUENCE 442 AA; 45690 MW; F64FA816E9CF2434 CRC64;  
 Query Match 8.0%; Score 147.5; DB 1; Length 442;  
 Best Local Similarity 25.6%; Pred. No. 0.0044;  
 Matches 80; Conservative 28; Mismatches 115; Indels 89; Gaps 14;  
 QY 14 LALGT--EMASAE-----CSKLYGO--CGK-----NMNG-----PTCCSG 47  
 DB 150 IASGTKMDLVLSEONLIDSKSYGNNCGEGLMTLGFYIINKGIDTSSYPYTAEDG 209  
 QY 48 STCKVSNDRYS-----QCLPSSSSGKSSBSAHKTTTA--AHKTTTAHKKTTTAPA 99  
 DB 210 KEKFKFTSNIGAOIVSYOVNTTSGSEASLOSASNNAPVSAVIDASNSFOLYESGIYYEPA 269  
 QY 100 KKT-----TYAKASTPSSSSSSGKSAVSGAGNGVTRRYMDCKKASCSMPGKA 152  
 DB 270 CTFQDLGHVLYVVGSGSSSSSSGSSSSSSSS-----STGKTSSSSSSKASSSSGKA 326  
 QY 153 NVSSPYKCNKDVTLSDSNAGSGNGNSYNCNDNPMAVNDNLAYGFAAAAISSGGE 212

Db 327 SSSS---SSGKTSAASTSGSGSGSGSQ-----SGQSTGSG 362  
Qy 213 SRMCCSFEFLTFTSTVAGKMKVQVNTGDDLGSSGTAHFDLMPGGGIVGIFNGCSQW 272  
Db 363 S-----CGTASAGS---ASAGSGSGSGSGSGSGS-----SGSGGAVFASGNY 401  
Qy 273 GAPNDGWSRYG 284  
Db 402 WIVNWSWGTSMG 413

## RESULT 15

GUNS TRIRE STANDARD; PRT; 242 AA.

AC P4317;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)  
DE (Cellulase V) (EG V).  
GN EGL5.  
OS Trichoderma reesei (Hypocrea jecorina).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Hypocreales; Hypocreaceae; Hypocrea.  
OX NCBI\_TaxID=51453;  
RN [1]  
RP SEQUENCE FROM N. A.  
RC STRAIN=CM9414 / Rut C-30;  
RX MEDLINE=95075308; PubMed=7984103;  
RA Saloheimo A., Henriksat B., Hoffren A.-M., Telemann O., Penttilä M.;  
RT "A novel, small endoglucanase gene, egl5, from Trichoderma reesei  
isolated by expression in yeast."  
Mol. Microbiol. 13:219-228(1994).  
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
linkages in cellulose.  
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL  
HYDROLASES).  
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; Z33381; CA83846.1; -.  
DR HSSP; P00725; ZCBH.  
DR InterPro; IPR000254; CBD\_fungal.  
DR InterPro; IPR000334; GH\_45.  
DR Pfam; PF00734; CBM\_1; 1.  
DR Pfam; PF02015; Glyco\_hydro\_45; 1.  
DR ProDom; PD001821; CBD\_fungal; 1.  
DR SMART; SM00236; FCB; 1.  
DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
DR PROSITE; PS50842; EXPANSIN\_EG45; 1.  
DR PROSITE; PS01140; GLYCOSYL\_HYDROL\_F45; 1.  
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 17  
FT CHAIN 18 242  
FT ACT\_SITE 27 242 ENDOGLUCANASE V.  
FT ACT\_SITE 134 134 NUCLEOPHILE (BY SIMILARITY).  
FT DOMAIN 18 182 PROTON DONOR (BY SIMILARITY).  
FT DOMAIN 183 205 CATALYTIC.  
FT DOMAIN 206 242 PRO/SER-RICH (LINKER).  
FT CARBOHYD 182 182 CELLULOSE-BINDING (BY SIMILARITY).  
FT DISULFID 213 230 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DISULFID 224 240 BY SIMILARITY.  
SO SEQUENCE 242 AA; 24411 MW; CC033FC51326C71D CRC64;

Query Match 8.0%; Score 147; DB 1; Length 242;  
Best Local Similarity 67.6%; Pred. No. 0.0027;

Matches 23; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 29 LVGQCGGKXNMNPTCCSGSSTCKVNSNDYYSQCLP 62  
Db 209 LVGQCGGAGWGTGPTTCQAPETCKVQNMYSQCLP 242

Search completed: June 18, 2003, 15:32:01  
Job time : 10.8656 secs